



AI技术赋能企业应用

上海信弘 庄健炯

2025 年 06月



目录

01

我们的业务

02

英伟达的企业
AI解决方案

目录

01

我们的业务

02

英伟达的企业
AI解决方案

公司介绍 - 上海信弘智能科技有限公司

上海信弘 (ZENTEK) 是 NVIDIA公司在中国的**精英级合作伙伴**。我们专注于NVIDIA数据中心解决方案。

智算推动通过AI和数字孪生技术的数字化转型，为**智算中心、教科研、能源、金融、工业、医疗、设计、媒体**等领域的客户提供基于GPU架构的软硬件解决方案。我们的服务包括从咨询规划、测试选型、到部署实施并提供后续运营维护。

我们总部位于上海徐汇区，在北京、深圳、成都、马来西亚和美国设有分支机构。

DPU & DOCA卓越中心

培训
服务交付伙伴

Elite 级别资质

Compute | Networking | NVIDIA Omniverse | NVIDIA Virtual Desktop | NVIDIA AI Enterprise



DPU & DOCA Excellent Center

NVIDIA Technologies
— DLI

NVIDIA Compute

NVIDIA Networking

NVIDIA Omniverse

NVIDIA vGPU

NVIDIA AI Enterprise



智算中心

教科研

能源

金融服务

工业制造

医疗与生命科学

建筑设计

媒体娱乐

ZENTEK

COPYRIGHT © 2018-2024 ZENTEK.ALL RIGHTS RESERVED

公司介绍 - 上海丽蟾科技有限公司

上海信弘控股



Lichan

上海丽蟾科技有限公司

英伟达 NCP

英伟达 CND

(China NCP Center)

专业提供 GPU 架构的智算云服务
以及端到端智算中心解决方案

智算中心解决方案

- 提供全栈式智算中心解决方案。
- 根据客户业务需求进行规划和设计。
- 实施先进的计算和可扩展性技术。
- 设计到运维的全面支持。
- 结合 NVIDIA DLI 和合作生态系统培养智算中心人才。

GPU 架构的智算云服务

- 提供基于 NVIDIA 技术的端到端 GPU 云服务，提供优质算力以满足客户在深度学习、科学计算等领域的需求。服务包括 GPU、网络 and 软件，如 NIMs、NVAIE 和 OVE，专为 AIGC 和 AI4S 需求设计。
- 云平台具备弹性扩展和收缩能力，满足不同工作负载的需求，包括训练和推理，例如大语言模型 (LLM) 和行业特定的模型和应用。
- 自研云平台支持多种应用场景，包括人工智能、大数据分析和高性能计算，满足客户多样化需求。



成都天府云数据中心 - 英伟达的灯塔工程



为西南地区客户提供智算中心，加速AI工作负载。

面向电信、大型语言模型供应商、独立软件供应商和AI科学服务供应商等客户的AI人才发展中心。

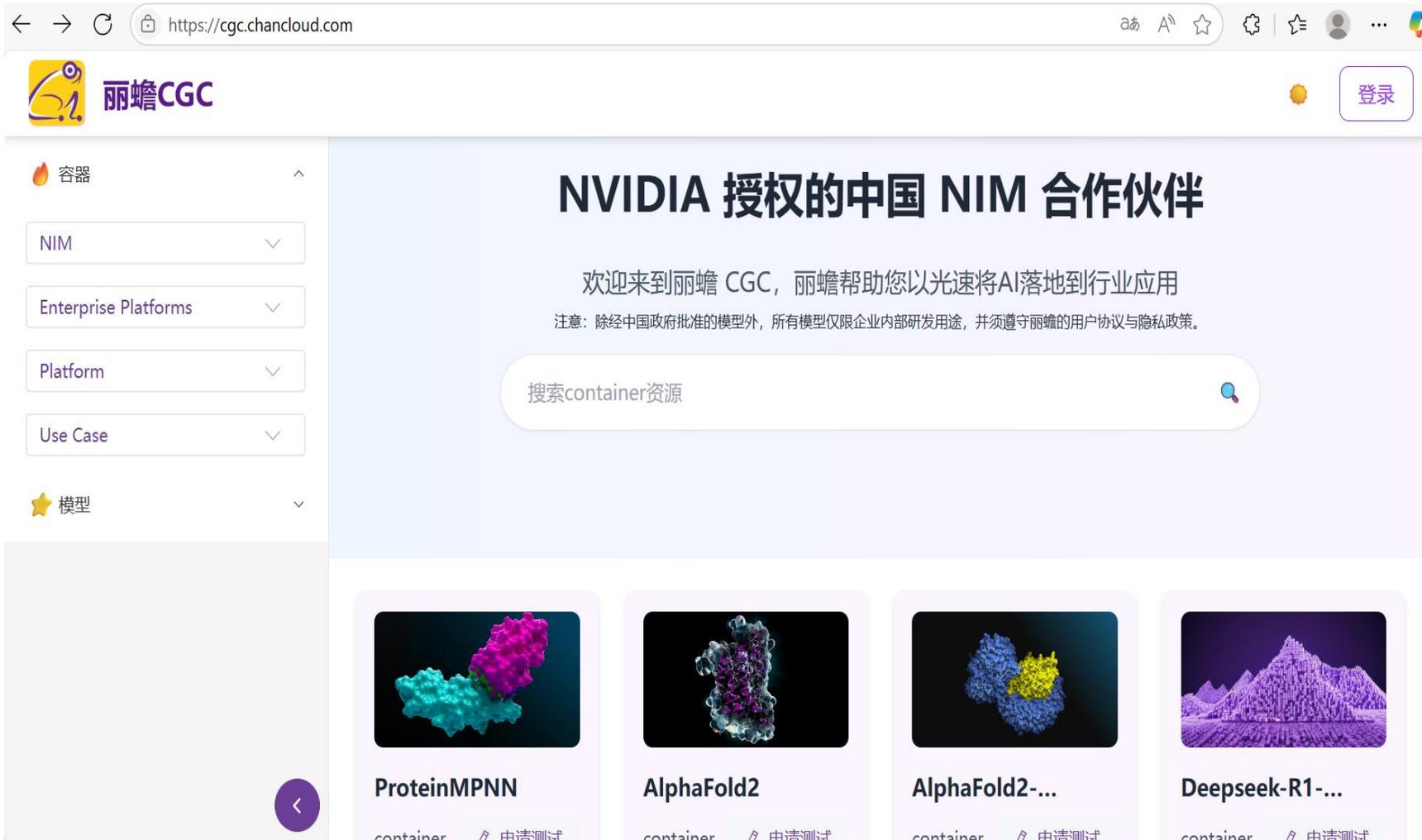
- 位于四川天府产业园
- 占地1600平方米，规划容纳400个机柜，提供9700P计算能力
- 目前已投入使用3200P
- IT电力容量为22MW（当前使用8MW）
- 目标PUE低于1.25



基础设施

- 硬件：NVIDIA GPU
- 硬件：NVIDIA InfiniBand
- 软件：NVIDIA Unified Fabric Manager (UFM)
- 软件：NVIDIA AI Enterprise
- 软件：自研资源管理平台Leaper

丽蟾CGC(China GPU Cloud)是目前唯一获得NVIDIA NGC官网跳转链接的中国平台 [\(https://cgc.chancloud.com/\)](https://cgc.chancloud.com/)



The screenshot shows the website interface for Chancloud. The browser address bar displays <https://cgc.chancloud.com>. The main header features the Chancloud logo and a '登录' (Login) button. A left sidebar contains navigation menus for '容器' (Containers), 'NIM', 'Enterprise Platforms', 'Platform', 'Use Case', and '模型' (Models). The central content area is titled 'NVIDIA 授权的中国 NIM 合作伙伴' (NVIDIA Authorized China NIM Partner). Below the title, a welcome message states: '欢迎来到丽蟾 CGC, 丽蟾帮助您以光速将AI落地到行业应用' (Welcome to Chancloud CGC, Chancloud helps you land AI in industry applications at the speed of light). A search bar is present with the text '搜索container资源' (Search container resources). At the bottom, four featured AI models are displayed: ProteinMPNN, AlphaFold2, AlphaFold2-..., and Deepseek-R1-..., each with a 3D visualization and a 'container' label.

- **丽蟾云 (chancloud.com)** 是面向人工智能产业打造的一体化AI赋能与服务平台, 全面支持从**算力服务、AI资源、AI训练与推理、AI应用部署**的完整流程, 助力 AI 各类场景的落地提速
- **NVIDIA 首批CND (China NGC Distribution)** 为中国境内用户提供一站式英伟达 (NVIDIA) 全球AI软件资源服务
- 目前已经同步上线**80+海外热门NVIDIA NIM服务**

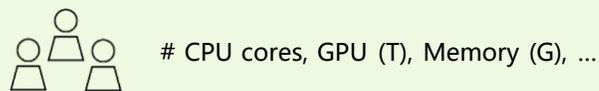
通过自主开发的Leaper资源管理软件可访问NVIDIA AI平台

资源管理

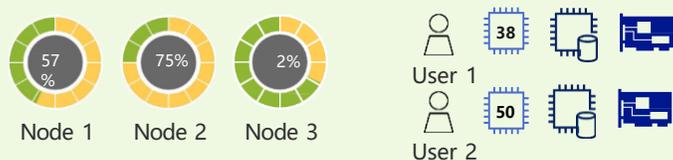
AI 平台

数字孪生 / Robotics 平台

资源申请



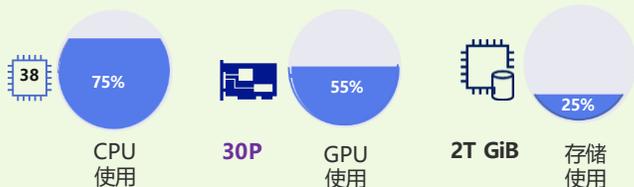
资源分配



资源监控



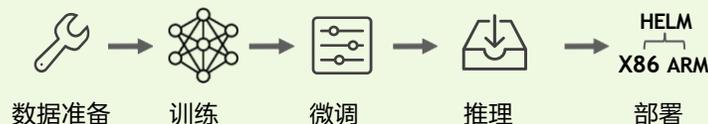
资源使用



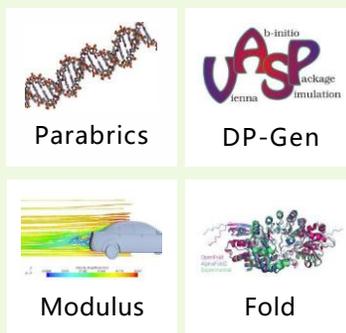
资源租赁



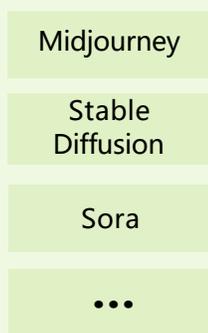
无代码平台



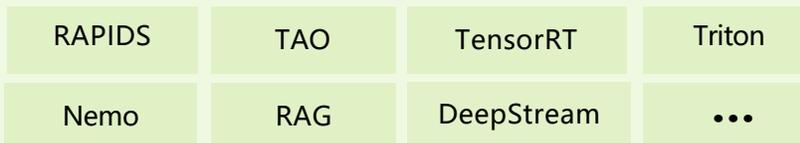
Alfor Science



AIGC Connector



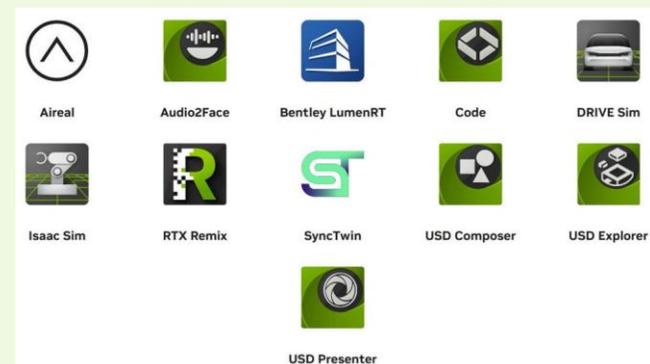
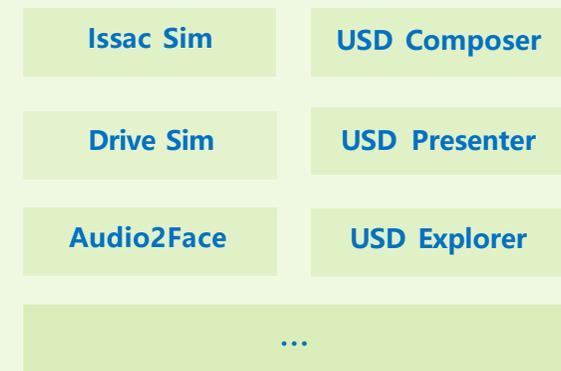
NVAIE 通用组件



NGC 资源池

可直接使用的库：容器, 预训练模型, SDKs, Helm charts, ...

NVIDIA Omniverse



- 已完成
- 进行中

NVIDIA 英伟达深度学习培训中心 (DLI) 课程

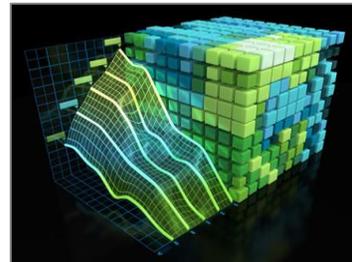
深度学习培训中心 (Deep Learning Institute) 是 NVIDIA 推出的深度学习教育和培训平台, 它旨在为世界各地的高校师生、科研人员和AI从业者提供高质量的 GPU 加速计算、数据科学、大语言模型训练等方面实训课程和虚拟实验环境, 帮

专业的AI 人才培养课程

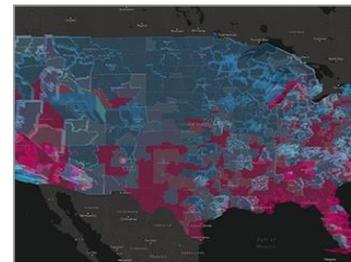
配套提供基于 NVIDIA 算力关键技术和AI应用领域的实战培训课程



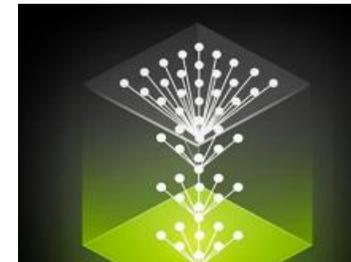
深度学习系列



加速计算系列



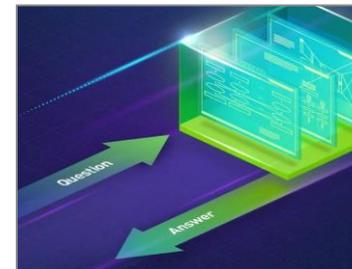
加速数据科学系列



大语言模型系列



AI - 网络安全



AI - 对话式 AI



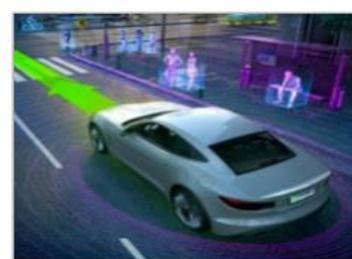
AI - 推荐系统



图形与仿真



异常检测



自动驾驶



机器人



医疗



行业课程

面向各特定行业有人工智能应用的从业者

多行业

使用 Omniverse Audio2Face 和 Riva 构建数字人

金融业

构建基于深度学习的反欺诈应用

制造业/工业

构建基于深度学习的工业检测应用
机器人训练与仿真

更多课程敬请期待

NVIDIA深度学习中心 | 认证课程

面向有人工智能应用需求的非专业从业者：
科学家 艺术家 企业家 律师 及面向市场的从业人员



深度学习必备技能系列

- 深度学习基础——理论与实践入门
- 构建基于 Transformer 的自然语言处理应用
- 数据并行 —— 用多 GPU 训练神经网络
- 模型并行 —— 构建和部署大型神经网络

加速计算系列

- 加速计算基础: CUDA C/C++
- 加速计算基础: CUDA Python
- 加速计算进阶: 用多 GPU 加速 CUDA C++ 应用
- 加速计算高阶: 将 CUDA C++应用扩展至多节点

生成式AI & 大语言模型系列

- 深度学习基础——理论与实践入门
- 构建基于 Transformer 的自然语言处理应用
- 构建基于扩散模型的生成式 AI 应用
- 构建基于大语言模型的应用

数据中心核心课程 | 原理·部署·运维

面向系统及网络管理人员

应用工具&管理平台

AI训练基础架构
GPU + 网络 + 存储

人工智能通识课

面向企/事业管理者
有人工智能应用需求的非专业从业者

人工智能的基本原理

最新行业技术与应用

人工智能行业趋势

英伟达生态的认知

探索“元宇宙”

更多课程敬请期待

线上课程

面向大众的低门槛视频普及课

定制课程

基于现有课程的定制

我们在以下四个领域持续为客户提供价值

人工智能、大数据和超性能运算平台

基于 NVIDIA 的核心产品，我们为客户提供人工智能、大数据产品与专业服务，包括设计，构建与运维



行业解决方案

为设计、教育科研、制造业、金融等行业的客户提供行业加速计算解决方案包含：数据加速，AI低代码平台，模型加速部署等

人工智能人才培养

- NVIDIA DLI 认证
- NVIDIA GPU 与网络知识分享与培训
- NVIDIA 软件的知识分享与培训 (OVE, NVAIE)



定制化开发

- GPU 资源管理平台
- 基于 NVIDIA CUDA 的开发
- 基于 NVIDIA DOCA 的开发
- Omniverse 二次开发

目录

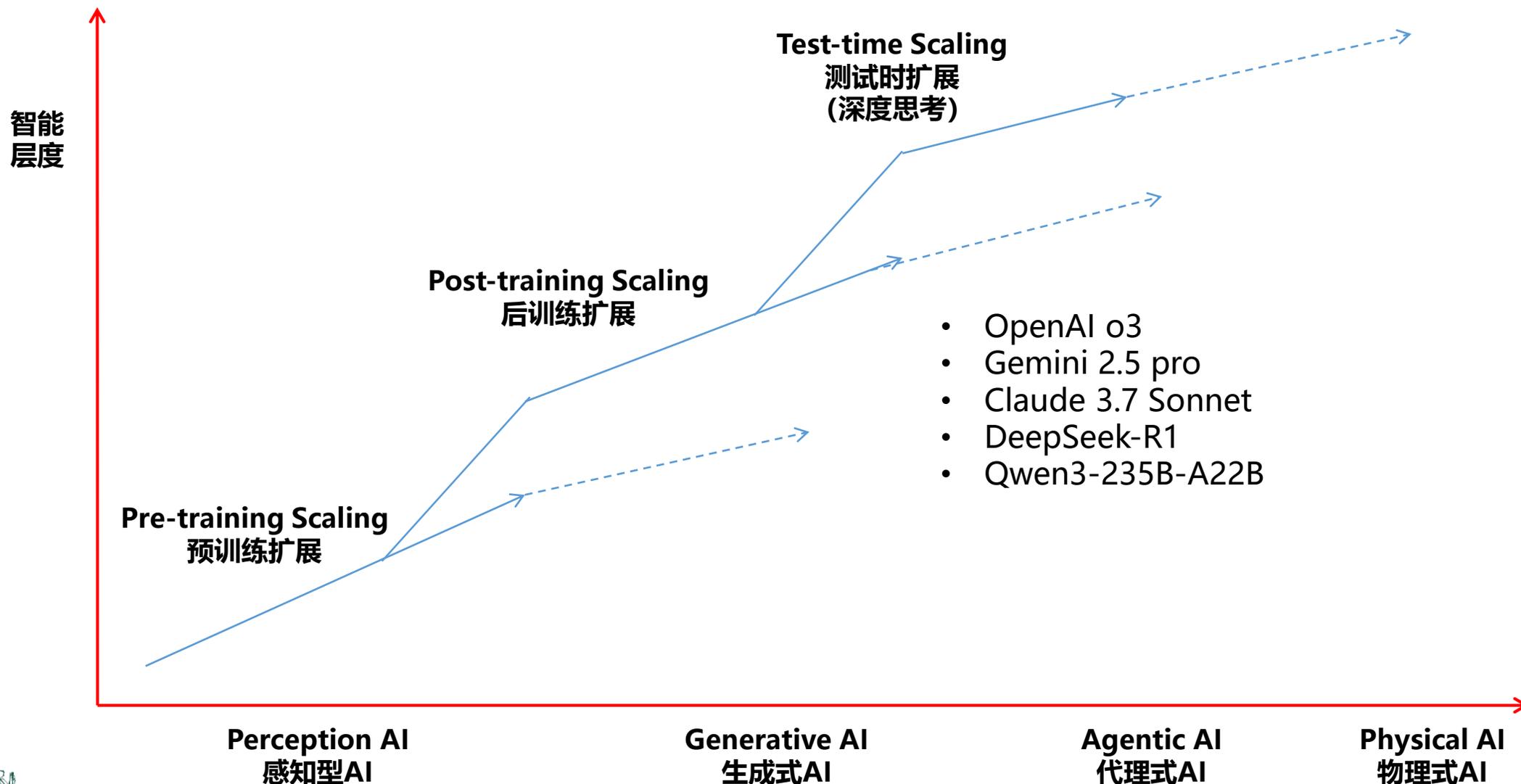
01

我们的业务

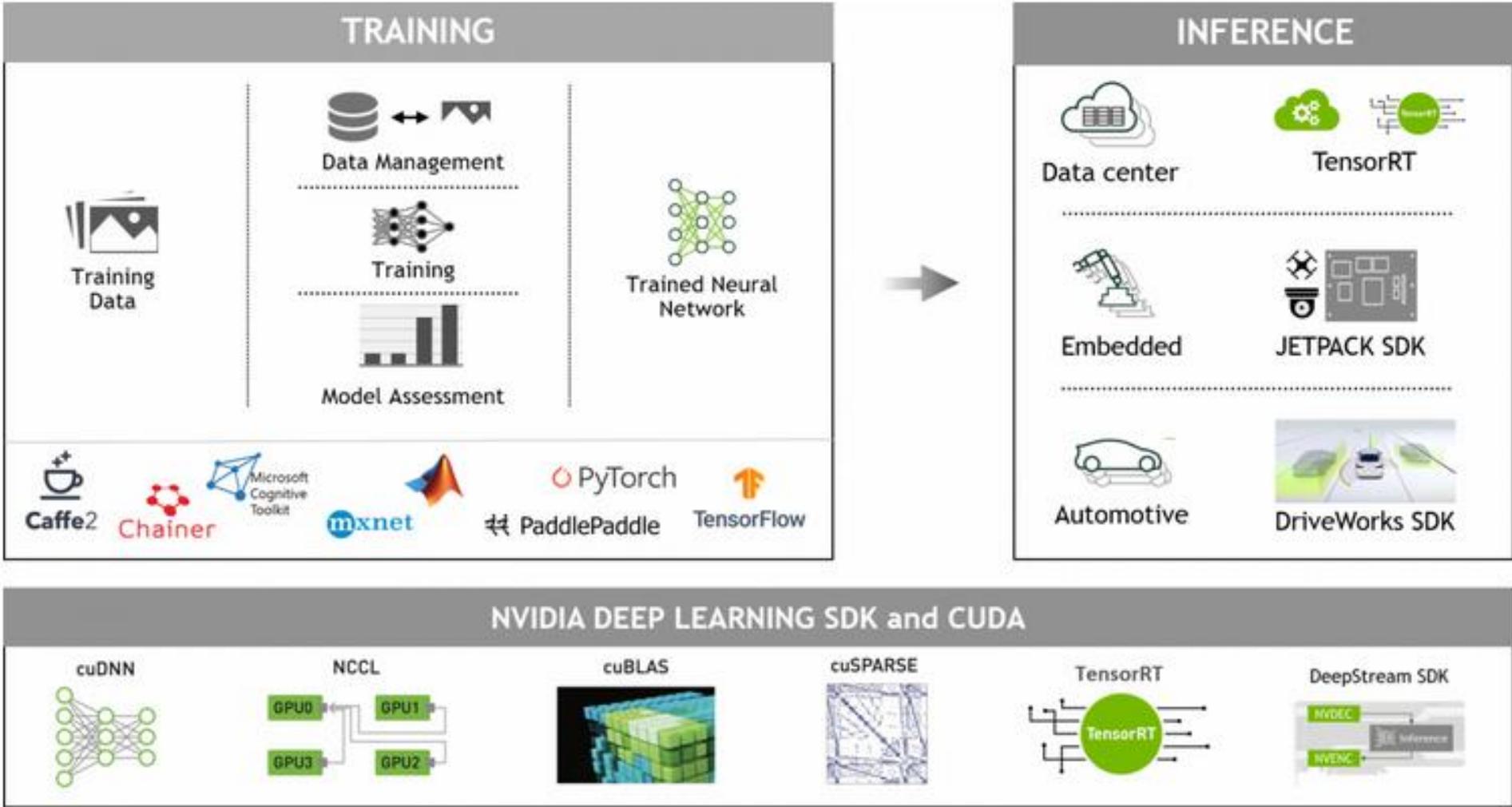
02

英伟达的企业
AI解决方案

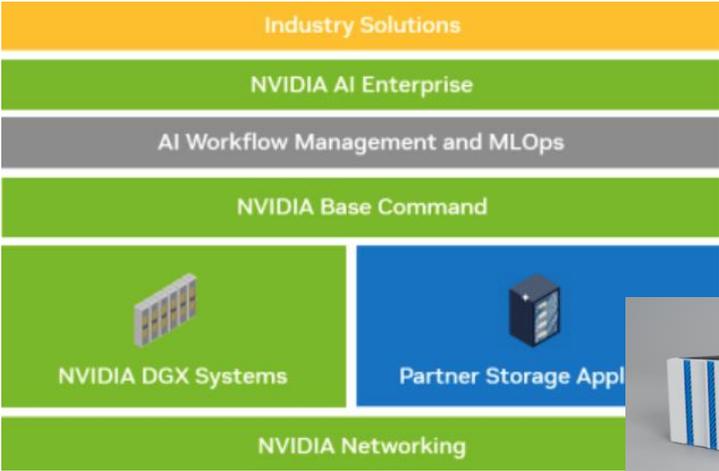
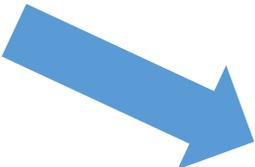
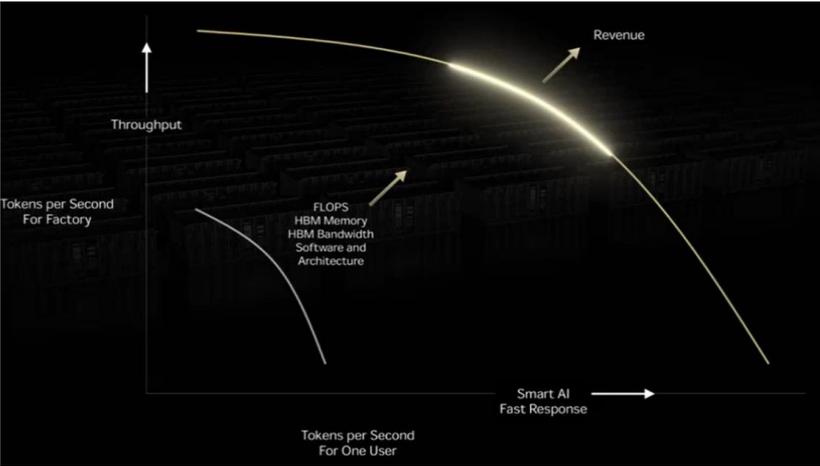
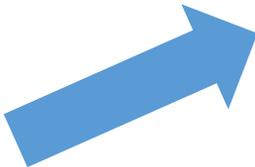
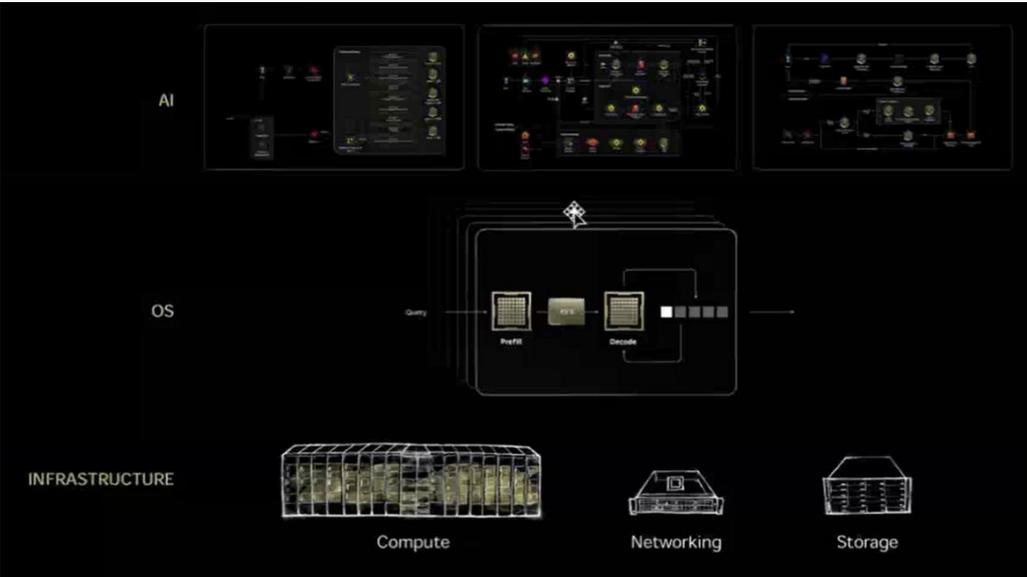
随着DeepSeek R1、OpenAI o3等推理模型的广泛应用，测试时(Test-time)的规模效应(Scaling Law)越发显得重要



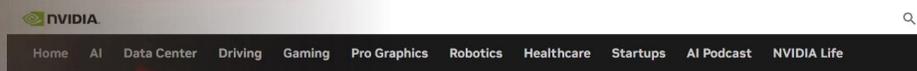
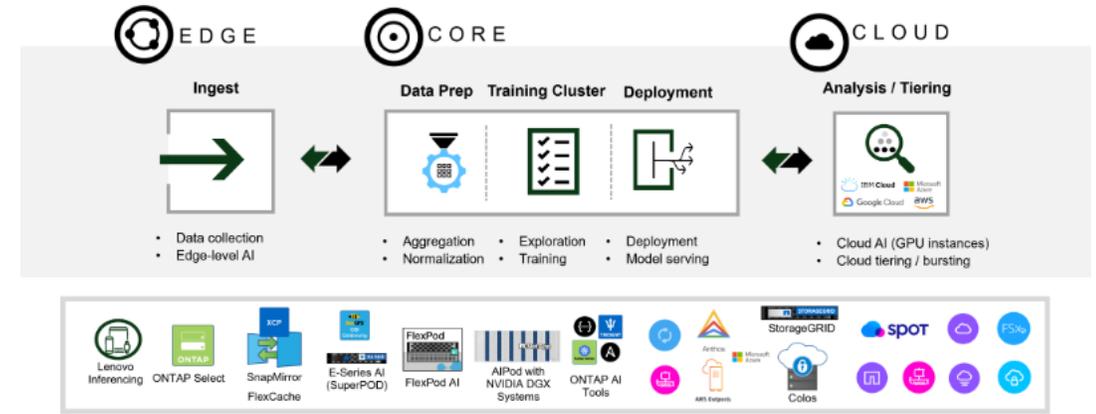
深度学习需要大规模的并行计算



推理型模型和代理式AI(Agentic AI)对企业IT基础架构提出新的要求



NetApp、Lenovo是NVIDIA全球战略合作伙伴

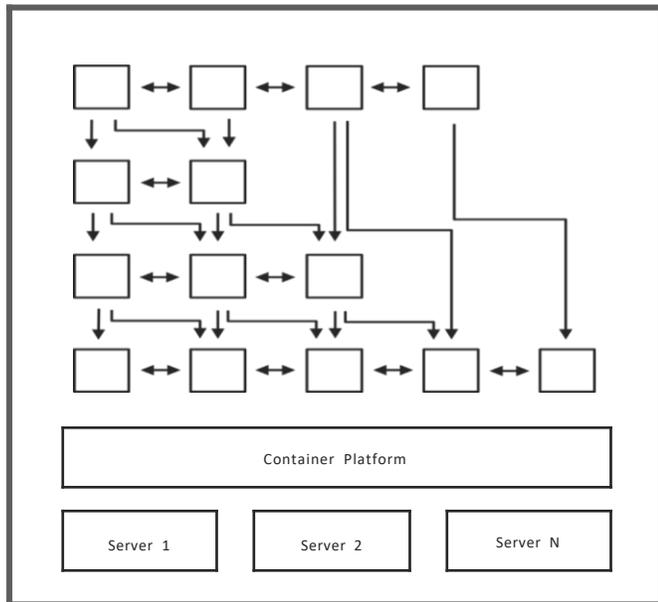


NVIDIA and NetApp Team to Help Businesses Accelerate AI

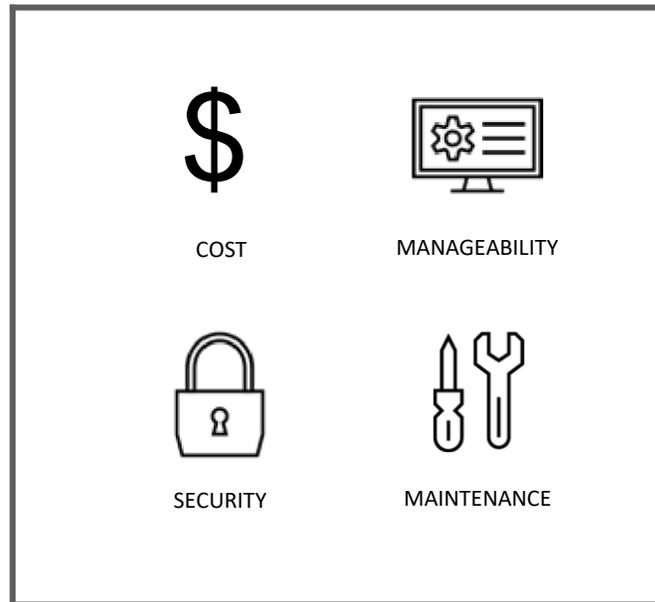
August 1, 2018 by [Jim McHugh](#)



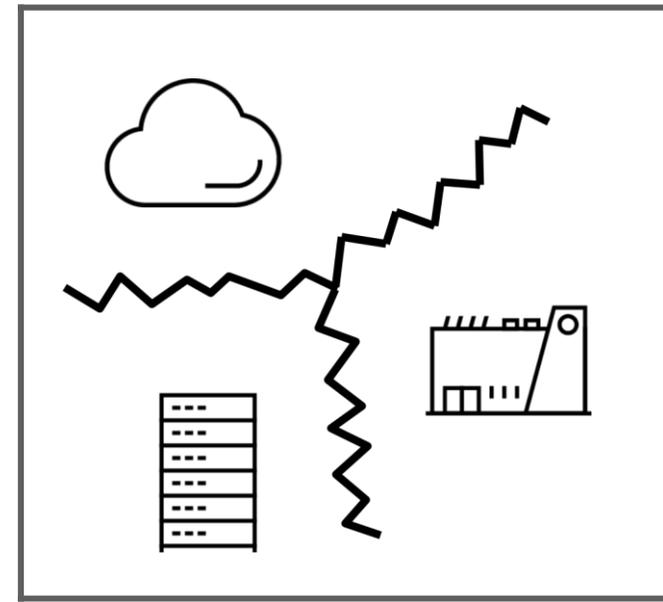
企业 AI 面临的挑战 - 不到 10% 的 AI 项目能够完成试点并投入生产*



部署复杂



生产就绪



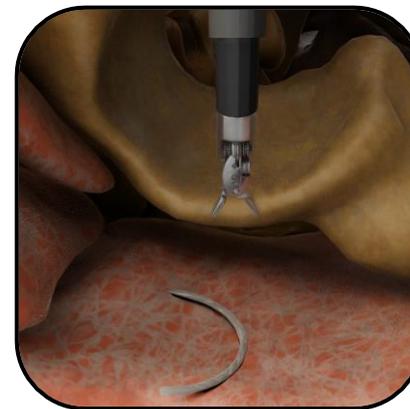
基础设施多样性

*Forbes.com。"生成式人工智能试点未能投入生产的原因"，2024 年 1 月 8 日。

英伟达的“三台计算机”架构



训练



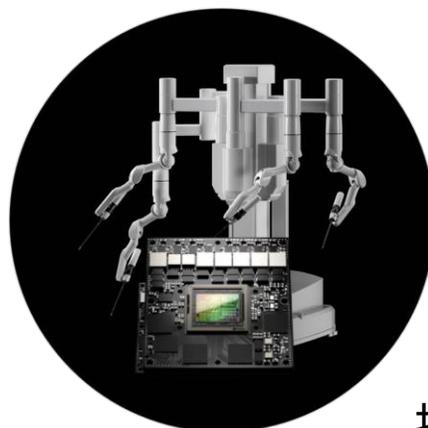
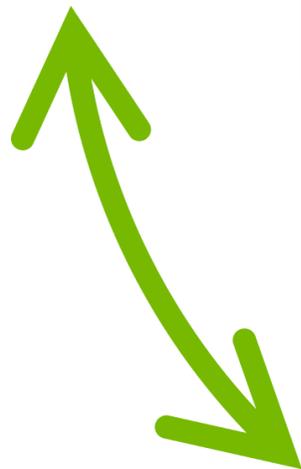
仿真



NVAIE

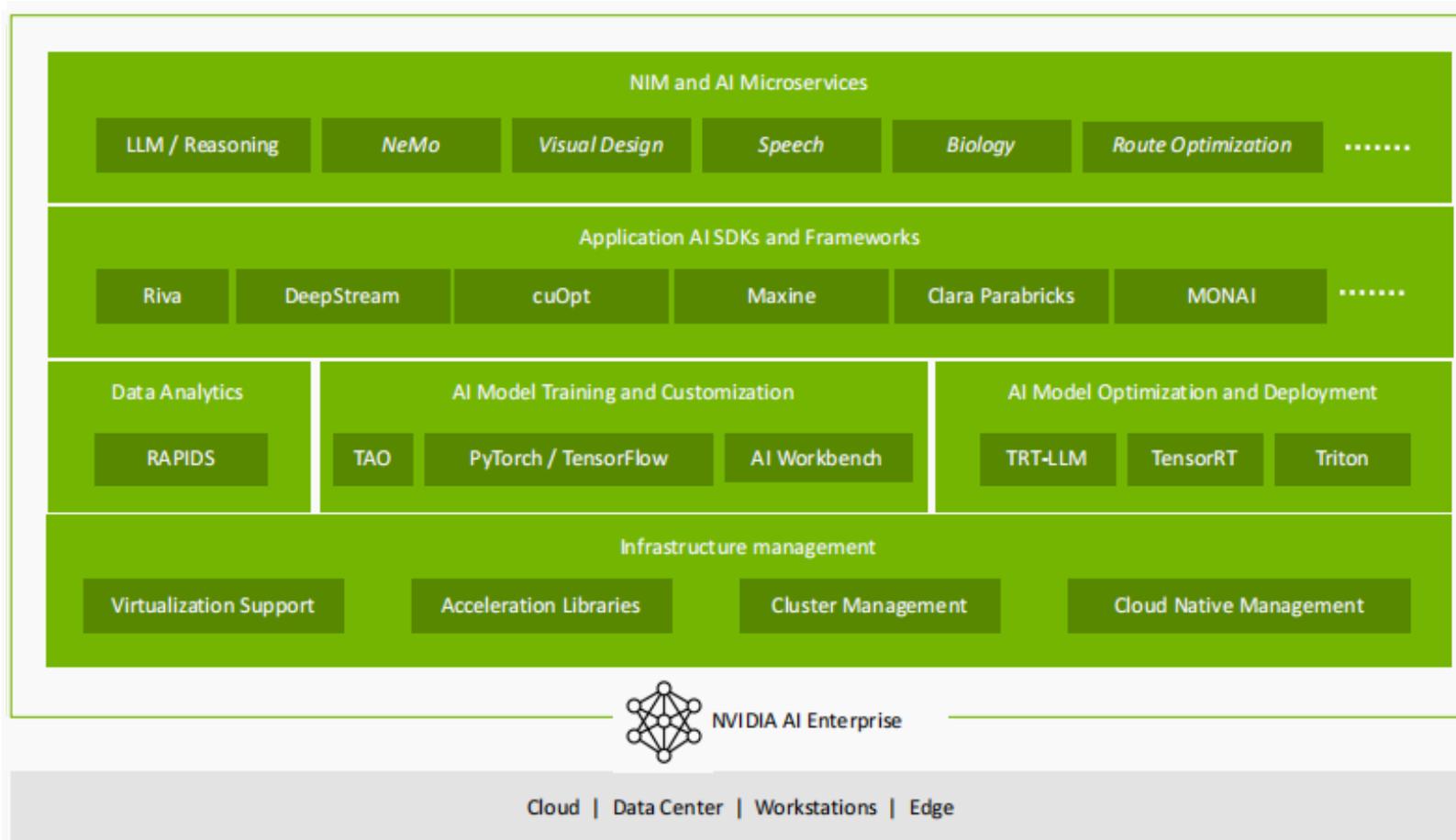


Omniverse



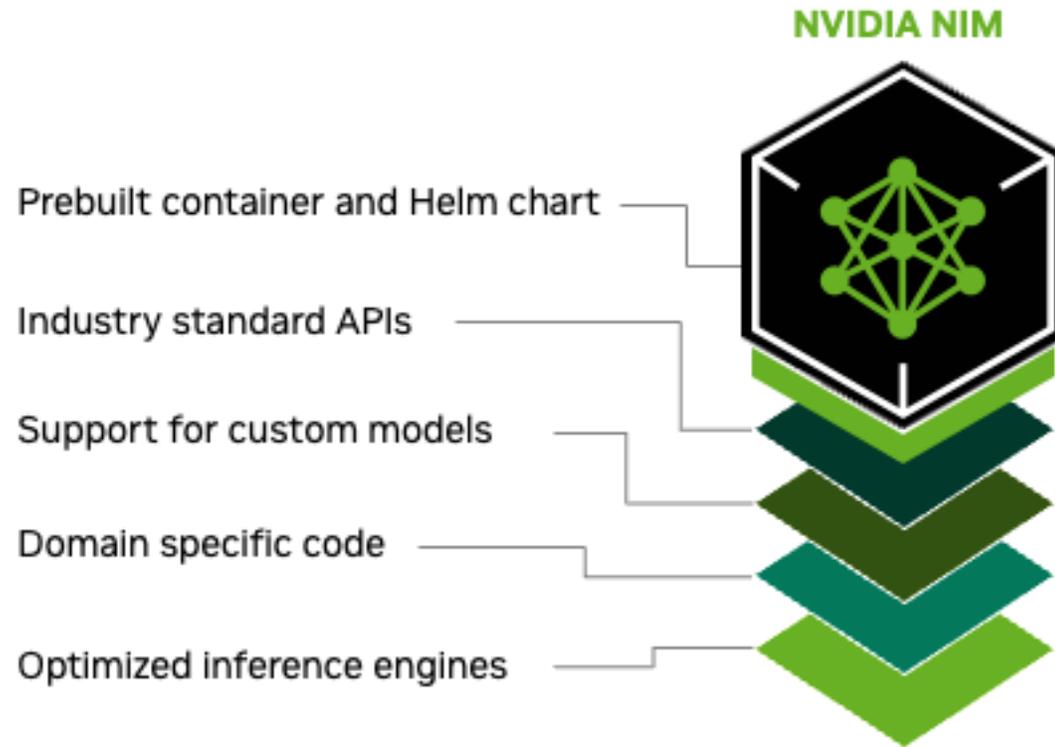
推理

NVIDIA AI Enterprise (NVAIE) - 面向AI的云原生平台



- 端到端的AI软件套件, 包括50+框架和预训练模型
- 云原生, 混合优化
- 部署在任何地方 - 云端和本地
- 提高AI模型准确性
- 安全且可扩展

用于生成式 AI 推理的 NVIDIA NIM 微服务



Deploy anywhere and maintain control of generative AI applications and data

Simplified development of AI application that can run in enterprise environments
Day 0 support for all generative AI models providing choice across the ecosystem

Improved TCO

Best accuracy

Enterprise software with feature branches, validation and support



DGX &
DGX
Cloud



DISCOVER

[Models](#)[Reasoning](#)[Vision](#)[Visual Design](#)[Retrieval](#)[Speech](#)[Biology](#)[Simulation](#)[Climate & Weather](#)[Safety & Moderation](#)

Discover

Experience Test-Time Scaling with DeepSeek

[Try Now](#)

State-of-the-art, high-efficiency LLM excelling in reasoning, math, and coding.

Most Popular Models [View All](#)

The leading open models built by the community, optimized and accelerated by NVIDIA's enterprise-ready inference runtime.



deepseek-ai
deepseek-r1



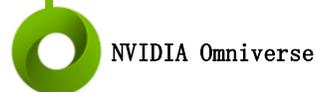
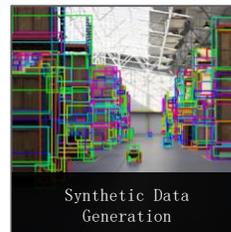
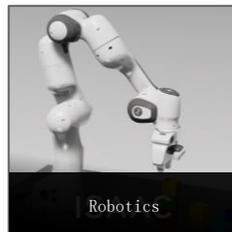
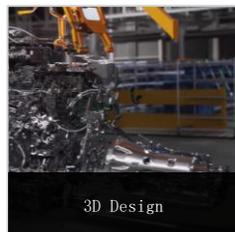
nvidia
cosmos-1.0-diffusion-7b



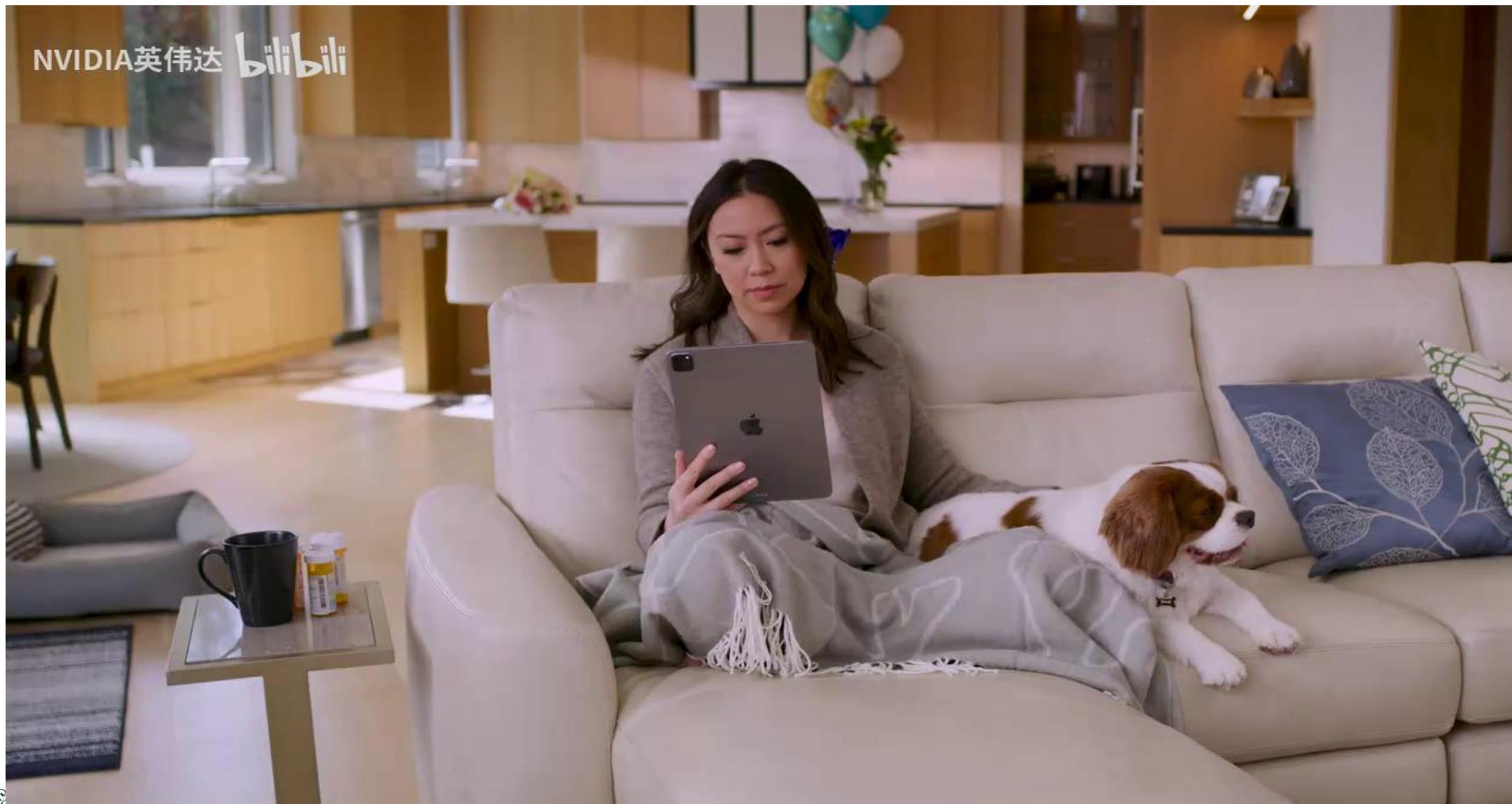
nvidia
cosmos-1.0-autoregressive-5b

NVIDIA Omniverse

连接、构建和运行行业数字孪生应用的云原生平台



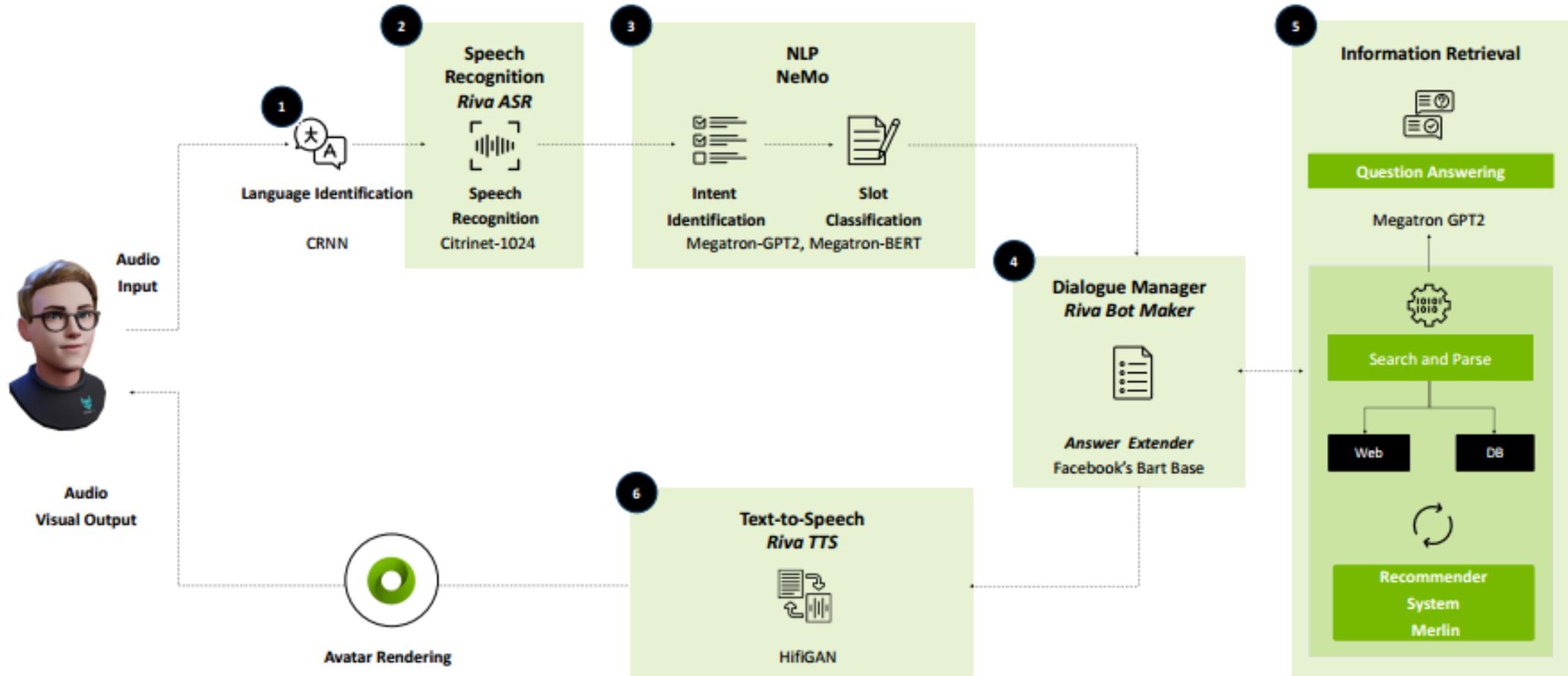
演示-用NVAIE和Omniverse实现的医疗数字人



NVIDIA英伟达 bilibili

NVIDIA Blueprint - 预构建、可定制的AI workflows参考实现

ASR-LLM-TTS-A2F





感谢观看

www.zentek.com.cn

COPYRIGHT © 2018-2024 ZENTEK. ALL RIGHTS RESERVED

